June 12th, 2017

Dimensionality reduction:

Traditionally dimensionality reduction is to create new features that are linear combinations of existing features. For example suppose have three dimensional data given below.

	f1	f2	f3
x0	3	2	1
x1	1	5	2
x2	10	1	10
x3	2	3	11

Let the first new feature be

newf1 = 10*f1 + .2*f2 + 3*f3 = $(10,0.2,3)^T$(f1,f2,f3)
newf2 = 2*f1 + 4*f2 + 6*f3 = $(2,4,6)^T$(f1,f2,f3)

	f1	f2	f3	=>	newf1	newf2
x0	3	2	1		33.4	(Similarly we make the newf2 features)
x1	1	5	2		17	
x2	10	1	10		..	
x3	2	3	11		..	

This is also called feature extraction. If we have a vector w then we can obtain a new feature by performing w^Txi for all xi.

Suppose each vector w was a random vector. This means each entry of the vector is randomly selected from a Gaussian with mean 0 and variance 1 or uniformly from the interval [-1,1]. It turns out that random dimensionality reduction performed in this way will preserve the distances between data points in the original feature form.

PCA will give us a vector w such that the projected data has the maximum variance. The projection of the data is simply

wTxi for all xi.

We can then plot the values of wTxi on a line. We could also look at the second principal component (called u) and plot its projection uTxi for all xi.